DOCUMENT RESUME

ED 121 825 TM 005 256

AUTHOR Smith, Philip G.
TITLE Structure of Evaluations in Education.
PUB DATE [Apr 76]
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Decision Making; *Effective Teaching; *Evaluation; *Evaluation Methods; Guidelines; *Research Design; Teacher Evaluation

ABSTRACT

In recent years there has developed a wide-spread recognition that there are three distinct structural or logical forms than an assessment of merit, value, or worth may take--evaluation of exemplified value, evaluation of instrumental value, and evaluation of contributory value. These forms provide a logical organization and structure for all of the various modes of evaluation that occur within the field of education, or elsewhere. This examination of the ways teaching is evaluated demonstrates how the three structural forms function in any assessment of merit or worth, whether it be grading student papers, selecting textbooks, or deciding to continue, modify, or discontinue a program. An adequate evaluation of any complex undertaking requres a design that makes explicit what structural form is to be used for each component part. Only then can it be known what standards or ideals and what ends of objectives need to be formulated and what organic wholes need to be envisioned. Then it is possible to know what kind of empirical data will be relevant to each component and what kind of expertise will be required for the judgments needed to bridge the gaps that cannot be spanned by exact measurement and calculation. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121825

STRUCTURE OF EVALUATIONS IN EDUCATION

by

Philip G. Smith

Indiana University

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A paper to be presented at
AERA San Francisco, 1976

TM005 256

During the second half of the twentieth century there has developed a widespread recognition that there are three distinct structural or logical forms that an assessment of merit, worth, or value may take. These three forms thus provide a logical organization and structure for all of the various modes of evaluation that occur within the field of education, or elsewhere.

There are still some terminological disagreements about the proper naming for these three basic structural forms. But the following names seem to me to be as good as any:

1. Evaluation of exemplified value.
2. Evaluation of instrumental value.
3. Evaluation of contributory value.

Some brief comments about each of these structural forms should be sufficient for this audience to see what kind of distinctions are made by this typology.

I

(1) To evaluate for exemplified value requires that one determine the presence or absence, or the degree of, certain properties or attributes that are held to be "good making" or "bad making" characteristics for the class or kind of thing that is under assessment. The method of inquiry is essentially empirical, ranging from carefully structured direct observations (as e.g. in judging diving or gymnastics) to complex forms of scientific analysis (as e.g. in assessing the nutritional value of some food product). The process by which a decision is reached is a matter of making accurate comparisons: What is being evaluated is compared to a set of standards or to an "ideal type" in order to determine to what extent the standards or ideals are exemplified in the actual case in point. Individual cases may then be graded in accordance with the standards, or they may be compared with one another and rank ordered, or compared with the distribution found in some normalized population and thus assigned to percentiles.

The corresponding non valuational research methods and decision processes are usually called "sorting and classifying." Specimens are examined and placed into various categories according to the properties or attributes they exhibit. From a

logical-structural point of view, the same kind of research and decision difficulties are encountered in evaluating for exemplified value that are encountered in sorting and classifying. A typical but non essential difference is that expert judgment rather than careful measurement and calculation frequently plays a more prominent role in assessing exemplified value than in sorting and classifying. This is because qualitative discriminations rather than measurements are more often called for in doing the research, and the decision or conclusion frequently requires an expert's sense of the relative importance of various combinations of characteristics rather than calculation in terms of a system of specified weights. But wherever such difficulties are encountered, whether in evaluation or in non valuational research and decision, there is no excuse for adopting arbitrary or pseudo-quantification techniques that gloss over the need for the careful training and experience required to develop competent observers and reliable expert judgment.

(2) To evaluate for instrumental value requires an assessment of the utility or effectiveness of something as a means toward or a causative agent in the production of some valuable end or product. The research methods are those of empirical-experimental inquiry. In the hard sciences such research frequently involves the kind of manipulation of factors or entities that permits strong inference to cause-effect relationships. In education we sometimes try to use various statistical techniques as a substitute for experimental manipulation and the results are frequently disappointing. Fortunately, however, a direct assessment of instrumental value is not nearly so critically important in educational evaluations as one might easily suppose. More will need to be said about this a little later.

(3) To assess the contributory value of something is to determine its merit or worth as a functioning part of some larger, valuable, organic whole. An adequate calculus for dealing with such organic or gestalt-like relationships is rather mind boggling--although not impossible in principle. Strangely enough, even ordinary citizens do sometimes engage in this form of evaluation with what at

least appears to be reasonable reliability and success. I suspect that the appropriate methods are more nearly those of the creative artist (or creative scientist) than those of "problem solving" as ordinarily conceived. In composing a painting, or a symphony, or a poem, for example, the artist considers a line, color, or arrangement of sounds in terms of how well it fits or "works", as they say, as an organic part of the larger whole that is being created. Do the lines or sounds make a contribution to the desired over-all effect while at the same time being in harmony with other contributing parts? Or, if these lines and sounds were to be incorporated, would they remain otiose or even distracting and of negative value?

Since in order to reach a conclusion about the contributory value of something it is necessary to determine how it would fit into a larger whole that, typically, must be imaginatively envisioned rather than actually experienced all at once, there are two techniques suggested by C. I. Lewis (1962) that enable one to reduce the problem to more manageable size. (a) The larger organic whole may be broken down into a number of smaller organic parts. The contributory value of the item in question may then be assessed in relation to one or more of these parts. For example, an ideal of architecture is that <u>all</u> parts must be subordinated to an over-all effect. Yet, typically, an architect does not attempt to devise or judge <u>every</u> detail directly in that relationship, but considers rather how a certain item will fit in relation to some lesser part, such as a room, an entrance way, or the like. In judging the contributory educational value of studying, say, a certain book, or topic, or of taking a certain course, we do not attempt to judge its value in relation to a <u>total</u> education, but rather in relation to some part of education, such as history, or literature.

(b) The second technique consists of distilling from the history of human experience, a number of <u>prima facie goods</u> whose contributory value may be taken for granted in the absence of reasonable doubt. For example, I don't have to conduct an evaluation to know that a good nights' rest, or a nourishing meal are

prima facie "good things" that make a contribution to my life-style and even to a life that is good on the whole. In the same way we know that learning to read and to write are important contributory parts of any education. We no longer need to construct arguments to justify them; indeed, we would need very good reasons to exclude them from any elementary school program.

These three structural forms provide the base on which any actual evaluation may be designed. Various enterprizes, such as schooling, manufacturing, or farming, have developed certain evaluative modes that reflect the kind of decisions that are critical in their operations. Each evaluative mode utilizes one or more of the structural forms.

In the field of education we rightly expect the evaluative modes to be similar to those of other professional practice, say medicine or law, rather than to those of manufacturing or farming. Unfortunately, this point is not always clear to some persons pushing for inappropriate forms of accountability in education. In any event, there appear to be some four or five principal modes of evaluating teaching and an examination of these may serve to illustrate how the structural forms may be utilized in the field of education.

II

Evaluating Teaching

A careful look at the expression "good teaching," as used by various persons in various contexts, and some analysis of the kinds of data that are supplied in support of claims of good teaching, in connection with, say, promotion dossiers at the university, provides an opportunity to display some of the types of evaluations appropriate in an educational context. "Good teaching" refers to any or all of the following:

1. Professional competence
2. Instructional effectiveness
3. Programmatic worth
4. Client satisfaction
5. Educational value

(1) Professional Competence. In any professional practice there are standards by which the conduct of the practitioner may be judged. This is to say that the standards constitute a list of good making characteristics and the practitioner's performance is judged as exemplifying (to some degree) the professional standards. This is the usual mode of evaluation used by professionals in defending against charges of malpractice. In medical or legal practice, for example, the critical question concerning professional competence is not whether the patient lived or the case was won, but whether the practice was professionally defensible. Were the appropriate procedures selected and carried through skillfully? Were there errors of either commission or omission that violate "standard practice"?

When teaching is viewed as a professional practice we see that questions of competence and skill may be raised in connection with any of the actions (Kerr & Soltis, 1974) involved in guiding and facilitating learning. Typically, some broad categories are used. For example, Thomas Green (1971) has suggested three categories or dimensions that seem useful from an evaluative standpoint. The teacher's work can be assessed for its competence on the substantive-logical dimension, its sensitivity and skillfulness on the pedagogical-psychological dimension, and its adequacy on the institutional-managerial dimension.

In the present state of the art we do not have well formulated, widely accepted, professional standards and procedures to serve as criteria for the professional mode of evaluation. Since it is this mode that is basic in the other professions, the absence of professional standards in education contributes to the image of teaching as being something less than a fully professional practice. When we attempt to evaluate teaching in the professional mode, we are forced, therefore, to rely upon judgments that can not be well explicated and it is no wonder that many teachers tend to be fearful of and to resent such evaluation. The answer, of course, is not to omit this mode of evaluation and concentrate upon less professional modes, but to get to work on developing the profession. The history of the other professions indicates that progress along this dimension is

more a matter of organizational posture and determination than of reaching some particular developmental stage in the theoretical-empirical knowledge base that supports professional practice.(Smith, 1974).

In the days of the normal schools a start was made toward the development of standardized procedures for teaching that were roughly comparable to, say, procedures in legal practice, such as preparing briefs, making motions, examining prospective jurors, cross examining witnesses, and the like. This development was made possible by the wide acceptance of a single theory of instruction based upon a comprehensive philosophical and psychological theory of mind, theory of knowledge, theory of learning, etc. known as Herbartianism. When teacher preparation moved to the universities, however, it was not surprising that both the theory and practice of the normal schools disintegrated. Penetrating criticisms, consideration of alternatives, creation of new hypotheses, is an essential part of the climate and orientation of an institution of higher education, in contrast to a trade school. In the university setting, the emphasis in teacher preparation shifted from training to education. Significant gains were made in both the liberal education and subject matter preparation of those preparing to teach in our public elementary and secondary schools.

Unfortunately, in this setting, our schools and colleges of education have not yet attained a truly professional posture. We do not provide programs for the preparation of practitioners in which a sound Arts College education is blended with the kind of theoretical study and clinical training that would produce graduates comparable to say engineering or law school graduates. Yet even among Arts College faculties there is a trend toward greater emphasis upon "good teaching," including professional competence. It is not uncommon, in connection with reappointment or promotion decisions, for the instructional materials that have been selected or developed by a candidate, to be examined and evaluated. Samples of his classroom performance may be observed and judged with respect to both the substative-logical and the pedagogical-psychological dimensions. And, of course,

data can be made available concerning the institutional or managerial aspects of his teaching, such as attendance, reporting, preparation of book orders and library reading lists, and the like. Evaluative judgements are made by colleagues who, presumably, have some expertise in these matters and although the criteria used in judging the performance (or performance products) are not explicitly formulated as professional standards, it is clear that the intent is to assess the teaching from the standpoint of professional competence and that the structural form of evaluation is that of the assessment of exemplified value.

In schools of education, this form of evaluation is frequently employed for formative purposes in both teaching laboratories and clinical situations. For example, in a contrived situation, sometimes known as micro-teaching, a prospective teacher conducts a short lesson with a small group of students. The performance is video-taped so that later it can be carefully studies by the trainee under the guidance of the "methods" or clinical professor. To what extent does the performance exemplify the principles and procedures that have been set for the trainee to master? To what extent does the performance exhibit the timing, smoothness, and authenticity that are the marks of skillfulness?

In some countries, initial certification is not granted until the candidate has demonstrated professional competence by performing skillfully in an actual classroom with a panel of judges observing. This practice, while perhaps largely ritualistic (not unlike the graduation recitals in Schools of Music), may be of value, nevertheless, in that it establishes this form of evaluation as a keystone of the profession.

If however, the professional mode were the only form of evaluation used, fear could arise that teaching might degenerate into a set of elegant rituals such as rain dancing. Afterall, the performance of a rain dancer could be judged for skillful execution of standard steps with appropriately integrated ad lib components that would lend authenticity to the total performance. But if one becomes concerned about the effectiveness of rain dancing, the corrective is not,

as some may suppose, to fund research for measuring how much rain follows various forms of dancing. The problem is, rather, a matter of designing and perfecting professional standards that are defensible in the light of theory and clinical practice. Any time an evaluation is made of exemplified value, there is an implicit claim that the criteria (or standards) used are defensible. In the case of teaching, this means that a performance that is good in terms of the standards, is a performance that is well calculated to guide and facilitate learning in such a situation. If in the light of theory and clinical practice, there are no good reasons for believing that one way of performing is any more promising than any other way, then there can be no standards and no objective evaluation in this form.

(2) Instructional Effectiveness. Empirical studies that undertake to assess teaching effectiveness must, of course, be designed in a manner that avoids post hoc ergo propter hoc reasoning. If we really had no other grounds for believing that a certain teaching procedure was well calculated to guide and facilitate learning in a certain type of situation, then the fact that learning occurred following the use of that procedure in an actual situation would have only the same evidentiary standing as the fact of rain following a rain dance. Since, however, even in the field of pedagogy, we are not totally devoid of either theory or clinical insight, it is possible to design ways of conducting instrumental evaluations of teaching, that, especially when used in conjunction with other modes of evaluation, are worthy of considerable confidence.

There are a number of problems that plague anyone who attempts thoughtfully to assess teaching effectiveness. If teaching is "to guide and facilitate learning" and "effective" means "the decisive production of intended effects," then a direct assessment requires empirical evidence that the intended learning did take place and that it was a result, consequence, or effect of the teaching in question. On the one hand are all of the problems concerning what observable behaviors can count as reliable evidence of the attainment of various types of

objectives; (Smith, 1975) on the other hand are the problems of demonstrating that whatever learning occurred did so as an effect of the teaching. At best, we can seldom do more than generate data within a design that makes it plausible to suppose that a given case of teaching was probably effective or ineffective.

Consequently, even though a direct evaluation of effectiveness calls for an assessment of the instrumental value of teaching, we are usually willing to settle for less direct evidence that can be obtained by the use of criterion referenced pre-intervention and post-intervention tests. And such tests, of course, are in the structural form of the assessment of exemplified value. Teaching is a system of instrumental actions designed (in the light of theory and clinical insights) to result in a number of specified student learnings. If pretest-posttest results indicate that the specified learning did occur, then, unless there are specific reasons for supposing that something other than the teaching was the causative agent, we attribute the results to the teaching. We thus establish a kind of prima facie claim for the instrumental value of the teaching.

At the university level there are two kinds of data frequently included in promotion dossiers as evidence of effective teaching: test scores and testimony. When an instructor teaches a section of a course designed to develop clearly demonstrable abilities (for example, basic courses in a foreign language or mathematics, and some courses in the physical sciences) it is not uncommon for all sections of the course to use a common departmental final examination. If, over a period of several semesters, students enrolled in Professor X's section of the course consistently score better on the final exam than students in the other sections, this suggests the hypothesis that Professor X is the more effective teacher. Provided there are no unrefuted alternative hypotheses that also account for these results, promotion committees are duly impressed, especially so if the instructors of the next course in the sequence testify that the students from Professor X's section typically seem better prepared than students from other sections.

A common criticism of an emphasis upon this way of evaluating teaching is that such an approach may encourage instructors to lower their sights, so to speak. Most teachers could easily increase their demonstrable effectiveness simply by concentrating all their efforts upon preparing their students to do well on some common or standardized test. Since such tests typically assess achievement of only a sampling of the performance and behaviorally indicated objectives for the course (to say nothing of more complex and long range educational objectives), (Smith, 1975) heavy handed effort to demonstrate effectiveness may actually decrease the value of the teaching. Some instructors may even encourage less able students to withdraw from the course or to transfer to another section and they may neglect to establish conditions that discourage cheating on the final examination. In principle, however, if teaching is the attempt to guide and facilitate learning, ultimately, there is no fully adequate substitute for an evaluation of the effectiveness of a particular attempt in a particular situation. On the other hand, such case by case evaluations are not really an adequate substitute for the theoretical and clinical insights needed to guide the attempt in the first place. It is virtually impossible to arrive at reliable empirical generalizations about the relations of teaching and learning by a series of ad hoc assessments of teaching effectiveness under actual classroom conditions. The more or less "theory-free" empirical data gathering practiced by proponents of the accountability movement in education is more than a little dangerous to the development of teaching as a professional practice. After all, he who takes credit for the sunshine may be held accountable for the rain.

At the university, since selection or preparation of texts and other materials are usually an integral part of teaching, an assessment of teaching effectiveness does not usually involve any effort to separate the effects of the materials from the effects of the style and behavior of the instructor. When evaluations are to be used for formative purposes, however, it is frequently important to gain at least some probability estimate of what part of the learning

can be attributed to what component of the complex of efforts that are held together by the term "teaching." Needless to say, the more variables that are identified in the teaching, the learning, and the setting (or conditions), the more complex the research design that is needed. While certain statistical treatments of the data may generate suggestions of possible cause-effect relations, what is more likely to be useful are suggestions generated by what Voltaire used to call le bon sens.

Experienced clinicians frequently have useful hunches about what has caused what, in even very complex situations. There has recently been some progress in the attempt to formulate what are essentially clinical rather than statistical approaches to the problem of determining what part of an observed variance can be attributed to various factors or variables operating in such complex situations. For example, Michael Scriven (1974) has suggested that the techniques of the detective, trying to find out "who done it," may be useful. Sometimes criminals develop distinctive ways of operating that leave tell-tale traces at the scene of a crime. Perhaps traces of some kind can be identified in observed results that suggest certain factors rather than others as the causative agents. Or, as Sherlock Holmes might have put it, careful observation of many small details permit one to make large inferences.

What is at issue here is the question of what kind of methodological restrictions and safeguards are most useful in situations in which the direct experimental manipulation and control of variables is either not feasible or else immoral. Remembering that what we are after are reasonably dependable results to be used by intelligent, sensible persons, we should avoid supposing that it is possible and desirable to develop research methods that are so uncontaminated by individual interpretations, hunches, and insights, that they are, so to speak, fool proof.(Levine, 1974) Only if we intend to employ fools as evaluators, do fool proof methods seem desirable.

Evaluations of teaching effectiveness may, of course, include more than an

answer to the question of whether or not or to what extent the intended results were actually produced. Questions concerning the expenditure of time, energy, and financial resources may also be considered. The terms "efficiency" and "cost-effectiveness" sound more like a manufacturing enterprize than an educational one yet, since schooling absorbs enormous sums of money, it is not surprising that questions about cost effectiveness arise. The problem is to formulate such questions in ways that are appropriate in various actual settings. A friend of mine in the School of Music once commented that today a good string quartet is neither more nor less efficient than a good quartet of a hundred years ago.

Whatever methods are used to assess, infer, or attribute instrumental value to the various methods and materials of teaching, whenever one examines means-ends, it is always appropriate to raise questions about the means themselves, apart from whatever instrumental value they may have. Such questions may be posed by using the other two structural forms of evaluation: (a) what is the exemplified value of the means? (b) what is the contributory value? Regardless of how effective a particular means may be, if it is judged to have negative exemplified or contributory value from an educational standpoint, we shall see that there is a prima facie case for rejecting it.

(3) Programmatic worth. When learning is studied as an effect of teaching there is no reason to stop with effectiveness. This is to say that the effects of teaching may be considerably more than the intended effects. A given case of teaching is said to have been effective when there is evidence that the students learned what the teacher intended for them to learn. But did the teaching produce what might be called "side-effects"? As teachers, we probably do both better and worse than we intend. While the teaching-pupiling activities are directed toward, say, American Literature, the students may make significant gains (or losses) in their understanding of, say, American History--to say nothing of skill in composition, sensitivity to language, and modification of attitudes about everything from schooling to morality. Good teaching ought to mean not only effective teaching but teaching that has good effects.

In order to judge the value of effects, a larger or more comprehensive normative framework is required. This is why, for example, when studies of effectiveness are limited to what is assessed by the final examination, the testimony of instructors of subsequent courses carries considerable weight. If they believe that the students from Professor X's section typically seem well prepared for the more advanced work, the meaning of "well prepared" includes more than was tested by the final exam. More specifically, while the final exam may have assessed only performance and behaviorally indicated objectives, subsequent instructors may include in their purview configurationally indicated dispositional objectives and clue indicated process objectives. (Smith, 1975) This is to say that when an instructor finds a student "well prepared," he typically means not only that the student has retained the knowledge and ability assessed by the earlier final examination, but also that he displays interest in and a favorable attitude toward the subject, and the inclination and ability to go about pupiling in a productive manner.

The effects of teaching are judged to be good to the extent they constitute a significant contributory part of the goals of the larger program. In this sense "good teaching" means teaching that is programmatically worthwhile. Assessment of programmatic worth is in the structural form of evaluation of contributory value. And this is true whether the effects have been determined or attributed by empirical research on actual cases of teaching, or are <u>anticipated effects,</u> projected on the basis of a professional evaluation of the teaching. Either way, the question to be answered is "What is the contributory value of these effects (i.e. of these learning experiences)?"

At the university, it is generally recognized that teachers, and their effects on the education of students, are not standardized, interchangeable parts of a constant, uniform program. When Professor Zieglar is replaced by Professor Zaglar (even though Zaglar is assigned to teach the same courses), a change occurs, not only in these courses, but in the overall effect of the program. Were this not

the case the situation would be viewed with alarm; it would be a signal that the program had become unduly atomistic, mechanical, and inorganic, rather than a coherent, synergistic, contribution to education.

In any attempt to compare the programmatic worth of the effects of the teaching of Professor Zieglar and Zaglar, it is judgment rather than calculation that is usually required. Since anything like a complete assessment of the contributory value of a teacher's work is too mind boggling to be undertaken, judgment typically is concerned with only the more obvious merits or deficiencies. Perhaps, for example, Professor Zieglar has a special flare for working with non majors. If an important goal of the department's program is to provide significant learning experiences for non majors, then Zieglar's contribution may be highly valued programmatically, even though, for example, his materials and presentations are judged, professionally, to be less well organized or less brilliant than Zaglar's. In spite of the difficulties involved in this mode of evaluation, such judgments are called for (especially in connection with tenure decisions). Even a rather truncated evaluation of programmatic worth serves as a useful counterbalance against the possible trivialization of instructional objectives that may occur by an over emphasis upon effectiveness.

Perhaps the most important use of an evaluation in this mode is as a safeguard against unintended, undesirable consequences of using certain means that may be, at least in the short run, quite effective. And the other side of this coin is that evaluation in this mode may uncover advantages of various materials or procedures that were not thought of when they were prepared or adopted.

One of the reasons why teaching frequently has unforseen effects is the differential impact upon the individual members of a class. The challenge that is optimum for Danna may be overwhelming for Dennis; materials that open new vistas for Paul may cause Paula to "turn off". Yet the answer to the problems caused by individual differences is surely not to make <u>all</u> instruction individualized by means of computerized programmed materials.

Human beings are social as well as solitary and an education that is suitable for beings who must endure this human condition must include learning to cope with this bipolar tension. One of the ways that the profession may hope to gain better understanding of this complex issue is by evaluations that strive toward an assessment of all significant effects upon all individuals of every major component in the program. Again, what is too mind boggling to undertake all at once may be chipped away by more limited studies of a restricted number of variables that clinical insight suggests may be critical in various actual situations. For example, what, if any, are the differential effects upon, say, reading comprehension and compositional ability, of mastering the objectives of a semester of geography through a series of programmed instructional modules versus mastering the same objectives through a more traditional classroom approach? Or, what, if any, are the differential effects upon slow learners, under achievers, emotionally immature students, versus various categories of more successful students.

In abstraction, one can multiply examples indefinately; in actual situations one may more profitably "work backwards", so to speak, from critical differences that cry out for explanation. In a given program to what may the success of the top twenty per cent of the students be attributed? Would case histories of these students reveal traces of some common or similar earlier learning experience that contributed in some catalitic or multiplier manner? Pedagogical research, in contrast to the traditional research of Educational Psychology, is needed to assess the programmatic worth of the various effects that may be attributed to a range of instructional materials and methods. In any professional practice, the research of the foundational disciplines may provide theoretical insights for the general guidance of practice but praxiological studies are needed for an evaluation of actual practice in specific situations.

(4) Client Satisfaction. The most common form of data collection, in connection with evaluations of teaching, is a survey of students' perceptions and satisfaction or dissatisfaction. A typical form includes two rather different

kinds of questions: in effect, students are asked about how they perceive the teacher's performance, and about how they feel about it. For example, do they perceive the teacher as knowledgeable and enthusiastic about his subject, are his lectures, materials, and class discussions well organized, are his explanations cogent and effective? In his teaching does he give ample opportunity for questions and does he encourage open and frank statement of difficulties and disagreements? Finally, do the students feel satisfied with the teacher as a teacher; would they welcome the opportunity to take another course with this instructor; would they recommend this teacher to others?

Some of these questions seek the students' perception of the professional competence of the teacher. Such questions may be the same or very similar to some of the questions posed in a professional evaluation of teaching. This may be why such surveys are, unfortunately, frequently called "Student Evaluations of Teaching," but a survey of even carefully considered preferences is not, strictly speaking, an evaluation.

Most sensible teachers would (as do other professionals) think it useful to learn how clients perceive what is going on and how they feel about it. Client satisfaction is a kind of good in its own right even where research fails to develop any significant correlation with professional competence, effectiveness, or programmatic worth. But the same professionals may not be enthusiastic about being evaluated by their clients.

This is not to say that just because a person is a student that he has no expertise in any of the dimensions appropriate to a professional evaluation of teaching. In fact, in the case of graduate students in schools of education, some of them may well be better qualified to evaluate the strategic-pedagogical dimension of a teaching performance than are many of the instructors at the University. Even such students, however, are not in a good position to evaluate teaching performance in a course in which they are enrolled for credit. This is so, first, because of their personal involvement, and second, because, presumably,

the reason for being enrolled in a particular course is that they have not yet attained the level of knowledge and understanding required for evaluation of the structural-substantive dimension of the course. Nevertheless, client satisfaction is one meaning of "good teaching."

The results of such surveys of students' perceptions and feelings may provide data that are useful in completing the kind of picture that is needed for an overall evaluation of teaching. For example, if explanations by the instructor that are viewed as quite adequate by a professional evaluation from the structural-substantive point of view are, nevertheless, perceived by students as confusing, then something is wrong. Of course, such perceptions do not tell us what is wrong. But armed with this information about client perceptions we are more likely to locate the difficulty.

Unfortunately, because it is much easier to formulate, administer, and tabulate the results of student opinion surveys than to devise and carry out an evaluation, there is a tendency, especially in higher education, to accept client satisfaction as a substitute for an evaluation of teaching. But this won't do. Afterall, practitioner satisfaction is also good in its own right. Yet we rightly are not content to have only so called "self-evaluation." When,however, an instructor feels dissatisfied about his own performance it is also a signal that something is wrong. But neither the perceptions and feelings of the instructor nor the perceptions and feelings of the students are adequate evidence for an evaluation of teaching.

(5) Educational Value: In a free society, if a teacher is found to be teaching what is false or twisted, or to be teaching by means of indoctrination, propaganda, and emotional appeals that discourage or even denegrate free and impartial inquiry, the teaching is judged to be professionally incompetent. But given a society in which professional standards were dominated by false doctrine, it is conceivable that a case of teaching might be evaluated as professionally competent, effective, and programmatically worthwhile, yet still be miseducative

rather than educative in its impact. (e.g. The Hitler Youth Movement) Consequently, a consideration of educational value is needed to complete the meaning of "good teaching". In our society, we are inclined to assume that a careful professional evaluation of teaching plus some consideration of programmatic worth will bring to light any cases of seriously miseducative teaching. But since all institutions, including schools, seem almost by nature to develop routines that tend primarily to insure orderly and stable operation, le bon sens suggests that periodic evaluations are needed to detect the amount of goal displacement that such routines may cause.

What may be at issue here is the rather complex topic of the meaning, in an educational setting, of the principles of democracy and justice. In spite of many competent and insightful analyses of this topic by John Dewey and other philosophers of education, there still appears to be widespread missunderstanding. In a free society the policies and practices of education (as an institution) ought to exemplify the same basic principles of democracy that we strive to exemplify in our political arrangements. But this does not mean that the schools should be operated as if they were primarily political rather than educational institutions. Abstract principles are exemplified by policy and practice in context. As we move from one context to another it won't do merely to transfer the practice; what is needed is to create practice that exemplifies the abstract principle in a way that is appropriate for the new context.

For example, if we confuse the democratic principles of the right and responsibility of participation with the particular policies and practice we have invented to exemplify this principle in our political affairs, then we may conclude that our schools are undemocratic unless students elect representatives to a school council, settle by majority vote what topics will be studied, or at the very least, select for themselves a curriculum from a wide range of "elective" courses or even modules within courses. But are such political forms of participation appropriate in an educational context?

Since such questions have been discussed elsewhere (Smith, 1964) we need not reconsider them here. For our present purpose the points to be noted are first, that in assessing the contributory educational value of an experience we should, as professor Peters has put it, pay attention to the manner as well as the matter. (1963) Second, when the manner violates such basic principles as respect for persons, then there is a prima facie case that the experience will be miseducative rather than educative. It seems likely that at least one reason why our schools have tended to swing back and forth between a miseducative stultifying regimentation and a miseducative foolish permissiveness, is the false assumption that there is a necessary conflict between freedom and authority, and between democracy and education. Yet each ought to be viewed as a contributory part of the other three. All together they constitute an organic structure within which a variety of life styles that are good on the whole may be constructed.

In addition to this check against miseducative experiences, an assessment of educational value provides the larger context in which even very complex programs may be evaluated. Some years ago, for example, studies were made of the extent to which "student values" underwent change during four years of college at various institutions. These studies were in the form of empirical research designed to study effects. What was really at issue, however, was the possibility of monitering some aspects of "incremental progress toward long range educational goals." (Smith, 1975) As we noted earlier, individual teacher accountability for the attainment of such goals does not make much sense. Yet surely, there is some responsibility here on the part of the profession and on the part of particular institutions.

It is not uncommon for institutions to assemble data on the "success" of their graduates and to claim credit for having made a significant contribution to whatever success has been documented. If, however, the central purpose of schooling is education, then a serious interest in good teaching ought to include some

efforts to assess effectiveness or success in relation to this purpose. From the standpoint of schooling, this is a means-end problem calling for an empirical research design that permits evaluation, in the instrumental mode, of the institution's programs. From the standpoint of education, the results of these programs (i.e. what has been learned) should be judged in terms of their contributory educational value.

It is probably inane to suppose that our great multiversities should or could provide the same kind of contribution to a student's value commitment and life style that seems to be provided by residence for four years at some of our better but small Liberal Arts Colleges. Nevertheless, any institution that is seriously trying to promote progress toward long range educational goals should make some effort to evaluate outcomes. Modern computer facilities at our large universities make it feasible to undertake very large scale, long term, follow-up studies of students and alumni. The kind of data that could be generated by such studies would surely be useful for diagnositc and formative purposes--including reformation of the data gathering process. At the very least, such an effort would symbolize a certain seriousness of purpose at a time when it may no longer be taken for granted that even an affluent society will continue indefinitely to provide ample financial support for educational institutions just on the basis of tradition.

One reason we are in trouble is that schooling has been sold in terms of its instrumental value. Correlations between the number of years of schooling and the amount of annual income have been used as arguments both for increased financial support of education and for discouraging individuals from "dropping out". But much stronger arguments could probably be constructed in terms of contributory value for both the society and the individual. The difficulty in constructing such arguments, however, stems from the fact that the Socratic dictum about the unexamined life not being worth living is hardly understandable, let alone convincing, to anyone living an unexamined life. Consequently, the

question "Why be educated?" as is the question "Why be moral?", is an indication of the novitiate state of the questioner. Such questions are, perhaps, best answered by the gentle but earnest advice, "Try it, eventually you'll like it."

In our society, however, we have presumably reached a state of culture in which we have attained a critical mass of educated persons capable of understanding the contributory value of education as an organic part of the good life. For such persons it is neither necessary nor appealing to construct instrumental arguments as a vindication of education. Nevertheless, it seems likely that, first, not all who are capable of understanding do, in fact, understand, and second, when educated persons become disenchanted with our schools, it is not the value of education as an abstract ideal that is at issue, but the policies and practices of actual institutions. This is why we need to work constantly at improving public understanding of education as the central purpose of schooling and why we need to undertake formative evaluation studies of the educational value of the effects that may sensibly be attributed to our schools.

III

It's a commonplace that education is pervaded with values and that most of the decisions involved in an educational enterprize are valuational decisions. It follows that if the decisions are to be made in a rational manner evaluations must be made. There is nothing new about educators making evaluations; they have always done it. What is new is that now we have a group of persons who are known as professional evaluators. Unfortunately, since many of these persons have been trained primarily as empirical researchers, they do not always see the connection between the kind of evaluations they are employed to conduct and the day to day evaluative decisions that teachers and administrators have always made as a part of their regular job. Indeed, some of these evaluators obviously don't really want to conduct an assessment of value; they would prefer to remain researchers. They prefer to define their own job as that of gathering data that may be useful

to decision makers. Under this approach, it is the decisions makers who make the evaluations, not the so called professional evaluators.

There is no doubt that research and measurement activities are an important part of evaluation. But as we have seen, empirical data can function in an evaluation only when placed in a structural context of defensible standards, ends, or wholes. If a decision maker has a clear evaluation design in mind, he may then employ researchers to generate the data he needs in order to make an evaluation. But when a person calls himself a professional evaluator, surely an employer has a right to expect that he is getting an expert on the structural forms and modes of evaluation as well as an expert on measurement and data generation.

Hopefully, this brief examination of the ways we evaluate teaching has shown how the three structural forms function in <u>any</u> assessment of merit or worth, whether it be grading student papers, selecting textbooks, or deciding to continue, modify, or discontinue a program. An adequate evaluation of any complex undertaking requires a design that makes explicit what structural form is to be used for each component part. Only then are we in a position to know what standards or ideals and what ends or objectives, need to be formulated and what organic wholes need to be envisioned. Then we are in a position to know what kind of empirical data will be relevant to each component and what kind of expertise will be required for the judgments needed to bridge the gaps that cannot be spanned by exact measurement and calculation.

References

Green, Thomas, The Activities of Teaching (N.Y. McGraw-Hill, 1971).

Kerr, Donna H. and Soltis, Jonas F. "Locating Teacher Competency: An Action Description of Teaching", Educational Theory, Vol. 24 # 1 Winter, 1974, pp. 3-16.

Levine, Murry, "Scientific Method and the Adversary Model: Some Preliminary Thoughts," pp. 661-677. American Psychologist, Sept., 1974.

Lewis, C.I. An Analysis of Knowledge and Valuation, (La Salle, Ill. Open Court, 1962).

Peters, Richard S. "Must an Educator Have an Aim?", pp. 83-95 Authority, Responsibility, and Education (London: George Allen and Unwin, 1963).

Scriven, Michael, "Maximizing the Power of Causal Investigations: The Modus Operandi Method," pp. 68-84 in James Popham (editor) Evaluation in Education (N.Y. McCutchan, 1974).

Smith, Philip G., Teacher Education at the Crossroads, The Bode H. Bode Centennial Memorial Lecture, (Columbus, Ohio, The Ohio State Univ. Press, 1974).

Smith, Philip G. "Instructional Objectives and Observable Behavior" Chapter 7 in Ralph A. Smith (editor) Regaining Educational Leadership. (N.Y. John Wiley, 1975).

Smith, Philip G. Philosophy of Education, Chapter 9. (N.Y. Harper and Row, 1964).